

## Notes to Figure 9:

MSV1 is the maximally specific rule version.

MGV1 and MGV2 are maximally general rule versions.

Only the rule patterns (left hand sides) are shown above.

All rules shown predict the same action: the appearance of a peak associated with atom "v" in the range 14.0 to 14.7 ppm. downfield from TMS.

The version space represented in Figure 9 above contains several hundred rule versions: the three versions shown plus all versions between these in the general-to-specific ordering. However, it can be represented simply by the two maximally general versions, MGV1 and MGV2, and the single maximally specific version, MSV1. The single most specific version contains every node and node attribute constraint consistent with all positive training instances. In this program the classes of positive and negative training instances are sets of molecules for which the indicated spectral peak does and does not appear. Thus, any rule version more specific than MSV1 cannot match every positive instance. Two general versions are required in this case since neither is "above" the other in the general-to-specific partial ordering. Any rule more general than either MGV1 or MGV2 will match some negative instance. Furthermore, any rule which is between these general and specific boundaries of the version space will match all current positive instances (by virtue of being more general than MSV1), and will match no current negative instances (by virtue of being more specific than MGV1 or MGV2).

### 3.1.3.3 Version Spaces and Rule Learning

Rather than select a single best rule version, the candidate elimination algorithm represents the space of all plausible rule versions, eliminating from consideration only those versions found to conflict with observed training instances. Thus, the candidate elimination approach separates the deductive step of determining which rule versions are plausible, from the inductive step of selecting a current-best-hypothesis. The algorithm is assured of finding all correct versions of the rule after all training data has been presented without the need to backtrack to reconsider previous training data or decisions.

In this example, RULEGEN was used to generate a set of plausible rules characterizing the CMR spectra of a set of

training molecules. For each rule, the associated evidence was given to a the candidate elimination routine which formed the version space for this evidence set. Subsequent data may be analyzed to modify the version space in a manner guaranteed to be consistent with the original data.

The candidate elimination algorithm operates on the maximally general and maximally specific sets representing the version space. The set of maximally general rule versions (MGV) is initialized to a single rule consisting of the most general possible rule subgraph (a single atom graph with no constrained node attributes), and the predicted shift range determined by RULEGEN. The set of maximally specific versions (MSV) is initialized to a rule which contains as its subgraph the entire molecule associated with the first observed positive instance. The initial version space represented by these extremal sets therefore contains all rules which match the first positive training instance (the most general possible rule, the very specific rule, and all intermediate rules).

The training instances are then considered one at a time. Each training instance is used to eliminate from the version space those rule versions which conflict with that instance. This is always accomplished by shifting the maximally specific and maximally general boundaries of the version space toward each other as shown in Figure 10.

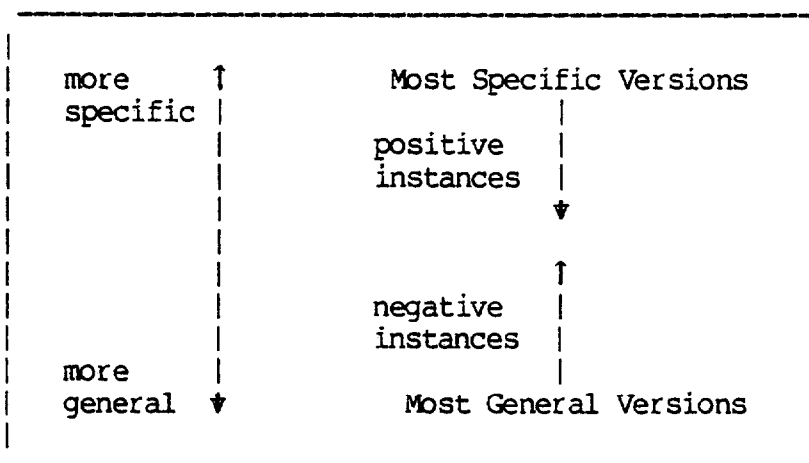


Figure 10. Effect of Positive and Negative Training Instances on Version Space Boundaries

Positive training instances force elements of MSV to become more general, whereas negative training instances force elements of MGV to become more specific. The maximally specific set can,

of course, never be replaced by a more specific set (nor the maximally general set by a more general one) since by definition, any version outside the current version space boundaries is inconsistent with previous training data. The action taken by the candidate elimination algorithm in updating the extremal sets is given below.

For negative training instances, each element of MGV which matches the instance must be replaced by a set of minimally more specific versions which do not match the instance. These new versions are obtained by adding constraints taken from elements in MSV in order to ensure that they remain more general than some MSV, and thus remain consistent with previous positive instances. Furthermore, each element of MSV which matches the negative training instance must be eliminated from the set (since it is already maximally specific, it cannot be replaced by a more specific version).

For positive training instances, any elements from MSV which do not match the new instance are replaced by a set of minimally more general elements which do match the instance. In order to ensure that these more general versions do not match past negative training instances, any which are not more specific than at least one element of MGV are eliminated. Elements from MGV which do not match the positive instance are eliminated.

After processing each training instance, the new maximally general and maximally specific sets will bound the space of all rules consistent with the observed data.

#### 3.1.4 Current Status and Future Work

The incremental learning ability for Meta-DENDRAL depicted above in Figure 8 is almost fully implemented, but as yet remains untested. Routines for defining and modifying rule version spaces are implemented, as well as the ability to filter out training data explained by a rule set. The major unimplemented portion of the incremental learning scheme is the process for merging new rules into the evolving rule set. The chief issue here is deciding when and how to chose among or merge new rules which are similar to existing rules. We expect to complete implementation and initial testing of the incremental learning ability during 1978.

Among issues associated with the version space approach which we expect to explore during the current grant period are the following:

- 1) Intelligent selection of new training data from examination of partial results.
- 2) Applying chemical plausibility

information to select a "best" rule version from among those contained in the version space.

3) The extension of current methods for dealing more completely with noisy and ambiguous training data.

4) The use of version spaces for merging similar rules.

### 3.2 New Capability To Emphasize Discriminatory Power

One important intended use of rules formed by Meta-DENDRAL is the prediction of mass spectra for use in structure elucidation: Predicted spectra for a set of candidate structures are compared by computer with the mass spectrum observed for an unknown compound, and on this basis the candidates are ranked according to the likelihood of their identity with the unknown. The ability of rules, in this context, to differentiate correctly among candidate hypotheses is called their "discriminatory power." Since the selection criteria previously used by Meta-DENDRAL during the various stages of rule formation did not necessarily correlate with high discriminatory power, it was decided to provide the program with the option of directly emphasizing discriminatory power during rule formation, in order to maximize the usefulness of the resulting rules for purposes of structure elucidation.

This addition to Meta-DENDRAL has now been designed and implemented. The general method employed by the new option is as follows. Observed mass spectra of the training molecules are analyzed prior to rule generation to determine how diagnostic the various observed peaks are, within the training set, of the molecules that show them. This information is then used during rule formation to compute a measure of discriminatory power for emerging rules. This measure is used, in combination with other criteria, to guide the search during rule generation, and to control the modification and selection of rules during the later phases of processing.

Preliminary testing of this new rule formation scheme on the monoketoandrostanes produced rules of considerably greater discriminatory power within that family than had been produced in earlier work with Meta-DENDRAL, even though the training set used was only half as large as that used earlier. This "discrimination option", now integrated with the new template-processing capability, is currently being further tested on a group of aromatic esters to determine whether the rules formed are consistent with what is known about the fragmentation modes of those molecules, and whether the rules have significant discriminatory power outside the training set used to form them.

### 3.3 Improved Ranking Capability

The program used within the Meta-DENDRAL framework to rank candidate structures has been improved in several ways.

A) The program now summarizes its own results and prints the summaries, thus eliminating much tedious manual analysis that previously was necessary. This makes possible a much more systematic and extensive investigation of scoring functions and their behavior than was previously possible.

B) A large number of new scoring functions have been made available, many of them specially designed for use with rules formed under the "discrimination option."

C) A new ranking method has been implemented as an option, with an eye toward improving the application of scoring functions in ranking. This new method eliminates duplicate explanations of peaks (which were previously permitted) in a principled way. The new method may be easier to justify theoretically, and yielded generally better ranking results than did the old method in tests performed with monoketoandrostanes. Further tests are planned with aromatic esters and marine sterols.

### 3.4 Data Selection Program

It is a commonplace of methodology that good inductive generalizations depend on variety in the data set. This is no less true in the context of rule formation by Meta-DENDRAL. Whether the goal is to discover rules of high generality or high discriminatory power, one's chances of achieving this goal [appear to] increase with increasing variety of training instances. This suggests that it would be useful to have a data selection program that would select the subset of the potential training molecules which has the greatest variety, in some appropriate and well-defined sense.

A preliminary version of such a program has been implemented, and experiments with it will soon be underway. The method employed has two steps: A.) Construction of an index of all the structurally different possible fragmentation environments permitted in the molecules of the set of potential training molecules (PT) by the "half order theory" of mass spectral fragmentation. B.) Construction of an n-sized subset of PT that contains nearly the largest number of different permitted fragmentation environments possible for a set of that size.

### 3.5 Feedback Loops

#### 3.5.1 Filtering with Respect to Existing Rules

The RULEGEN program is capable of accepting previously defined rules as a means of filtering the evidence obtained from INTSUM before the evidence is used for rule formation. As well as providing a convenient and natural feedback mechanism for the program, this facility also allows rules obtained from other sources to be used to reduce the space which the program must examine to find rules for a given set of data. In this manner, the program is able to focus attention on evidence which is not already explained by any of the rules which it is given.

A problem with this approach arises from the fact that the spectral evidence may often be the result of more than one fragmentation. Yet the filtering mechanism assumes that any evidence which supports a rule is completely accounted for by that rule. Tests are in progress to determine the limitations of this approach.

### 3.6 Program Improvements

#### 3.6.1 Defining Rules with EDITSTRUC

In addition to the programs which produce rules from the spectral data, other programs have been developed to allow a user to define a set of rules manually. Like the rules produced by RULEGEN and RULEMOD, these are rules of structure fragmentation which are expressed in terms of molecular subgraph descriptions. The programs for manual rule definition provide a simple yet useful language for the description of these rules. A principle part of this language is the EDITSTRUC language, developed for CONGEN. This allows us to take advantage of the advanced structure manipulation capabilities which are a part of the EDITSTRUC package.

The ability to create rules manually should be particularly useful in conjunction with the rule filtering mechanism of RULEGEN mentioned previously. This provides the chemist with a natural means of describing obvious rules which the program can eliminate from consideration before focusing on the remaining unexplained evidence.

#### 3.6.2 Stability Rules in INTSUM and RULEGEN

The programs have been generalized to allow the analysis of the mass spectral data from the point of view of determining

rules about stable bonds, i.e., lack of fragmentation in a molecule as well as fragmentation. Just as peaks are evidence of fragmentation in a structure, absence of peaks is evidence that certain fragmentations have not occurred.

The programs are now capable of examining the original data from either point of view and proposing rules of behavior of the molecules from that point of view. Further work remains to be done to carry this generality through the processing performed in RULEMOD and then in conducting experiments to determine the usefulness of stability analysis.

### 3.6.3 Expanded Template Space

Originally, the subgraph descriptions in the rules produced by the RULEGEN program were restricted by requiring that the internal connection patterns of the subgraphs had to be completely specified. In other words, for each of the interior nodes in the subgraph, the complete set of neighbors had to be specified. This restriction excluded ruleforms which seemed to be both plausible and desirable, so the program was changed to eliminate the restriction.

In terms of the mechanism used by the program to search the space, implementation of this change meant removing the restriction on the subgraph matching templates that the neighbors property be required at all but the outer levels of a template. This allows the program to find rules in which the internal connection patterns of the chemical subgraphs are only partially specified.

For example; it is now possible to express a rule such as "break any bond which is 2 bonds away from an oxygen atom". Such a rule could not be expressed previously without identifying whether the nodes between the oxygen atom and the break were secondary, tertiary, or quaternary.

### 3.6.4 Small LISP and Program Efficiency

Increased size and complexity of the Meta-DENDRAL software has resulted in increasing efforts aimed at making the programs more efficient and understandable. All the programs which are part of the meta-DENDRAL system are now capable of running in the environment of "small LISP". This makes considerably more memory space available to the chemist for the data structures, thus making possible the solution of significantly larger problems than were possible in the standard LISP environment.

### 3.6.5 Help Facilities

As the programs have increased in complexity and usefulness, we have had to face problems of documentation and explanation of the programs to its users. Text explanations of the various aspects of the programs must be provided, and kept up to date, to allow others to use the system. It is also important that the text descriptions of the programs be available to the programs themselves to be used during program execution to provide on-line guidance to the user concerning the use of the programs.

Text descriptions of the programs must be closely associated with the programs themselves to insure that program changes are reflected accurately in changes in the text which describes them. Yet text explanations must be incorporated into the programs so as not to take up space which should be available during program execution to be used for producing results. Attempt has been made to resolve these sometimes conflicting goals through the use of the comment facilities of LISP, and through the generation of programs and conventions for programming which allow program documentation and explanations to be incorporated into the programs as comments in the appropriate places. There are then programs which have access to this information to produce documents and on-line explanations about the programs.

## 4 COLLABORATIVE RESEARCH

### 4.1 CONGEN Users

Dr. Peter Gund of Merck, Sharpe and Dohme Laboratories contacted us for a current CONGEN manual and Guest login information. He now feels that he has analytical problems which would lend themselves well to checking with CONGEN.

Professor Richard E. Moore of the University of Hawaii visited Stanford and was provided with a CONGEN demonstration on a problem relating to his own marine sterol work. We discussed system access and Tymnet node availability with him. He plans to return in the near future with another problem, and then consider the possibility of requesting access.

Dr. Jean-Claude Braekman of the University of Brussels travels across Brussels to use a terminal at the offices of the Belgium Chemical Society, in order to access CONGEN on SUMEX. Dr. Braekman uses the mail facilities to remain in contact with Prof. Djerassi's research group.



Dr. Martin Huber, a postdoctoral fellow in Professor Wipke's SECS group has been starting work in an area which was related to the graph theoretic basis for CONGEN. In an effort to encourage cross-fertilization of ideas, we encouraged and arranged a meeting between him and several of the DENDRAL project members. The resulting discussion, at the least, provided Dr. Huber with suggestions and information for further study. Likewise, DENDRAL was able to obtain a better idea of similarities in research interests between the two groups. We are currently pursuing several problems in graph theory concerning analysis of molecular structures. These problems arose directly from this meeting and concurrent discussions with Prof. Wipke.

During the special symposium at the San Francisco ACS meeting in the fall of 1976 which Ms. Suzanne Johnson helped to organize and chair, members of the DENDRAL group provided on-line demonstrations of CONGEN during the "hands-on" session. At this time Professor Kurt Mislow of Princeton University expressed interest in using the program. Later, we provided him with Guest access information and answers to his questions concerning terminals and other useful programs available to chemists on various commercial networks. As a result of this effort, Professor Mislow has used CONGEN and has been considering its use as a teaching aid. He wrote us this past spring to enquire whether Guest access to CONGEN might be possible for his friend Professor Weiss, head of the Department of Chemistry at Northeastern University. We subsequently provided Professor Weiss with the information necessary to access CONGEN on a trial basis.

In November 1976, Dr. Stan Lang of Lederle Labs' Infectious Disease Research Section, requested access to CONGEN. After being provided with the appropriate information and initial help, he encouraged Dr. Leon Goldman to request access also, and to request information on obtaining a copy of the teletype DRAW program used to draw CONGEN structures on teletypes. A recent phone conversation with Dr. Babu Venkataraghavan, a new member of the research group at Lederle, indicated that the TTY DRAW program was being used quite successfully. Also interested in the possibility of support for graphics terminals, Dr. Venkataraghavan called to discuss the problem in terms of Omnigraph, which they already have on their PDP-10. We have exported a complete copy of all the DRAW program files, including ample data files, to Dr. Venkataraghavan and are currently in contact with him on implementation questions.

A further example of cooperation between DENDRAL and Professor Wipke's group concerns the sharing of graphics programs. DENDRAL obtained the Fortran sources for programs created by the SECS group to do molecular modelling and structure display on the DEC GT40. Wanting to interface these programs to CONGEN, but not wanting to limit CONGEN graphics to one terminal

type, DENDRAL personnel modified the program to use the Omnigraph graphics package available on SUMEX. Glenn Ouchi of the SECS project, has become familiar with the relationship of the graphics in CONGEN to the Modeller's graphics. SECS has become aware of the desirability of supporting additional terminal types for graphics output, and will be investigating Omnigraph applications to this area.

One of the students who used CONGEN in Prof. Djerassi's molecular structure elucidation course introduced the program to a graduate student of Professor E.J. Eisenbraun's (Oklahoma State University). Professor Eisenbraun is a well known marine natural products chemist. He has requested Guest access information, and appropriate materials were provided in spring of 1977. Professor Eisenbraun subsequently visited Stanford and got a personal demonstration of CONGEN.

We have been in contact with Dr. Karl Kuhlman, a chemist and PROPHET user at SRI International. We have arranged for a group of DENDRAL chemists to get together with the SRI group for exchange demonstrations: CONGEN for PROPHET, and discussion of similar problem areas with visiting PROPHET representatives.

Dr. David Pensak of Dupont in Wilmington, Delaware originally started out as a CONGEN Guest user. In return, he contributed a good deal of knowledge concerning evaluation and use of molecular modelling programs. At the current time he is beginning to build a research group in computer applications in chemistry, and views SUMEX/DENDRAL somewhat as a resource from which to obtain knowledge of hardware, software and people.

Dr. Milton Levenberg of Abbott Laboratories first expressed interest in CONGEN at an ACS meeting two years ago. He was given an account and appropriate information at that time. He had used OMNIGRAPH to develop a program to display and plot mass spectra, which he gladly provided to us. That program now provides a means for chemists to obtain a plot of their spectra which have been obtained on mass spectrometers which are not yet equipped with automatic computer output.

When Kent Morrill was a graduate student in chemistry he developed an interest in CONGEN and various of the Meta-DENDRAL programs. When he left recently for a job with Tennessee Eastman, he requested Tymnet login information to take with him. As a result of his interest, Dr. Gary Santee of Eastman Kodak in Rochester requested information for Guest access to CONGEN. Kodak may also be in the process of forming a computer applications in chemistry group, and once again, we seem to be viewed as a potential information resource in this type of effort.

Dr. Gretchen Schwenzer was a postdoctoral fellow with DENDRAL. When she left Stanford for a job at Monsanto, it was

with the idea of taking part in helping to develop a computer applications in chemistry group. She too views SUMEX as an information and know-how resource. To that end, we have had several phone calls and terminal links from her concerning graphics, terminals, modelling programs and text editors. She is interested in obtaining several copies of documentation preparation programs either developed or supported at SUMEX.

Dr. Robert Shapiro of New York University came to visit Stanford in September of 1977 to learn to use CONGEN. He spent a week in residence to discuss structure elucidation problems relating to nucleic acids and their interactions with other substances. We are also pursuing ideas on the automated analysis of UV spectra of such compounds, based on empirical rules derived from study of known systems.

In November of 1976, Dr. Henry Stoklosa of Ciba-Geigy approached one of the members of the DENDRAL project for trial use of INTSUM. During a subsequent visit to Stanford, we introduced him to CONGEN and its use. We have been keeping him up to date on recent developments because he indicated that CONGEN is beginning to have more and more use to him in the analytical task of evaluating additive bonding in polymeric materials.

Dr. Geza Szonyi of Polaroid corporation was one of the original persons to enquire about SUMEX/CONGEN access as a result of the "invitations for use" which were included as a part of early journal articles. He has recently requested trial access to CONGEN. Phone conversations indicate that his group is evaluating computer systems which will offer them the greatest latitude in applying computers to their work in various fields of chemistry and related data management. Once again, DENDRAL is viewed as a potential knowledge source.

Drs. D. Williams and R. McGrew from the Midland, Michigan site of Dow Chemical came to visit Stanford and receive an introduction to CONGEN. They were given a CONGEN demonstration, and as a result, requested a copy of the teletype DRAW portion of the program, which we sent to them. This brings to five the number of sites which are now using the teletype DRAW program in some fashion. Also included are: Lederle Labs in New York, (Dr. Babu Venkataraghavan); Dept. of Computer Science at SUNY, (Dr. Dave Larson); Dept. of Chemistry, Arizona State Univ., (Prof. Morton Munk); Dept. of Chemistry, Miyagi Institute, (Prof. Hidetsugu Abe); and Cambridge University, (Neil Gray).

#### 4.2 Marine Natural Products

#### 4.2.1 Mass Spectral File Search System

An attempt was made to obtain mass spectra for all marine sterols reported in the literature (Appendix A). The old mass spectral files were scanned and pertinent sterol mass spectra were digitized (a file of non marine sterol mass spectra were also acquired from the older files as a supplement to the marine file) (see Appendix B. Marine sterol researchers were requested to send samples of specific sterols which they reported or sterol mixtures known to contain the requested sterol (see Appendix B. In a few cases sterols were isolated from crude extracts of organisms known to contain the sterols. The high resolution GC-MS spectra of the available sterols were recorded using a Hewlett Packard 7610A gas chromatograph equipped with a 10' X 2 mm "U" shaped column (3 per cent Poly S-179 on gas chrom Q or 3 per cent OV-17 on gas chrom Q (column temp. 260 degrees C) and interfaced with a Varian Mat 711 double focussing mass spectrometer (equipped with a Watson-Biemann dual stage separator, an all glass inlet system and a PDP-11/145 computer for data acquisition). High resolution spectra were recorded for subsequent fragmentation analysis by the application of data interpretation and summary programs, e.g., INTSUM, and to facilitate handling of the data for construction of the searchable files. Within the framework of the available data acquisition and reduction systems, the rapid analysis scheme has been tested, and the advantages and limitations are the subject of the following section.

The spectra of 52 marine sterols were compiled in a computer searchable format. The spectra, which are essential to have available for careful comparison following the search report, have been plotted, and the plausible or established interpretations of the higher molecular m/e peaks have been indicated on the spectra. Spectral interpretations have been coded in Fig. 8 in a series of 32 symbols which have been appropriately marked on the spectra of each sterol in Appendix C which is the file of marine sterol spectra constructed in our laboratory. Attached is a list of investigators who reported and received copies of this file. This summary of proposed fragmentation rules is acting as a preliminary guide in the INTSUM evaluation.

The SEARCH program was used to match every spectrum in the file (Appendix C) to every other spectrum to gain an indication of how all the spectra rank to one another in terms of the similarity index described previously (Table V). A rank of 999 indicates a positive identification; therefore, each spectrum when compared against itself results in a rank of 999. Ranking values below 500 indicate positive nonidentity and are not recorded. Ranking values approaching 750 indicate a possible match is not ranking higher due to variations in spectrometer operating conditions. Table V displays a number of interesting results. First, several separate sterols rank at the identity

rank, that is, they have mass spectra which are similar enough to be basically indistinguishable:

Sterols 15 and 18 Appendix A: this indicates that mass spectrometry cannot distinguish between slightly different side chain alkylation patterns in some cases. This agrees with the similar evaluations in the literature.

Sterols 68 and 71: this indicates that mass spectrometry cannot distinguish between side chain double bond geometrical isomers (E and Z) in this case.

Sterols 90 and 80: these are again sterols with slightly different patterns of side chain alkylation.

See pp. 88a-c for Table V.

LIBRARY SEARCH REPORT FOR EXPERIMENT SEARCHING  
52 SPECTRA IN MARINE AGAINST THEMSELVES

STEROL		STEROLS MATCHED	
#	RANK	STEROL	
1	999	87 999	0 274 ANDROST-5-EN-3BETA-OL
2	999	92 999	0 300 PREGNA-5,20-DIEN-3BETA-OL
3	999	99 999	0 300 PREGNA-5,17(20)Z-DIEN-3BETA-OL
4	999	55 999	0 302 PREG-5-FN-3BETA-OL
5	999	90 999	0 314 23,24-DINOR-CHOLA-5,20-DIEN-3BETA-OL
	547	52 0	0 412 24-ETHYLCHOLESTA-5,24(28)Z-DIEN-3BETA-OL
	554	41 999	0 426 (24Z)-24-PROPYLIDENECHOLESTA-5-FN-3BETA-
6	999	107 999	0 316 23,24-DINOR-CHOL-5-EN-3BETA-OL
7	999	101 999	0 318 5ALPHA-23,24-DINOR-CHOLAN-3BETA-OL
8	999	91 999	330 330 24-NOR-CHOL-5-EN-3BETA-OL
9	999	99 982	0 370 24-NOR-CHOLESTA-5,22E-DIEN-3BETA-OL
	508	60 999	0 370 5ALPHA-24-NOR-CHOLESTA-7,22F-DIEN-3BETA-
13	999	109 999	0 372 24-NOR-CHOLEST-5-EN-3BETA-OL
10	999	118 999	0 370 5ALPHA-24-NOR-CHOLESTA-7,22E-DIEN-3BETA-
	576	68 999	0 398 5ALPHA-24-METHYLCHOLESTA-7,22E-3BETA-OL
	555	55 984	0 370 24-NOR-CHOLESTA-5,22E-DIEN-3BETA-OL
	504	56 999	0 384 CHOLESTA-5,24-DIEN-3BETA-OL
14	999	124 999	0 382 CHOLESTA-5-EN-23-YN-3BETA-OL
15	999	114 999	0 384 (24S)-27-NOR-24-METHYLCHOLESTA-5,22E-DIE
	800	76 0	0 384 CHOLESTA-5,22E-DIEN-3BETA-OL
	612	68 999	0 384 CHOLESTA-5,24-DIEN-3BETA-OL
	518	55 999	0 398 24-METHYLCHOLESTA-5,22E-DIEN-3BETA-OL
16	999	89 999	0 384 (24S)-5ALPHA-27-NOR-CHOLESTA-7,22E-DIEN-
17	999	113 0	0 386 (24S)-5ALPHA-27-NOR-24-METHYLCHOLEST-22E
18	999	95 0	0 384 CHOLESTA-5,22E-DIEN-3BETA-OL
	745	85 999	0 384 (24S)-27-NOR-24-METHYLCHOLESTA-5,22F-DIE
	594	66 999	0 384 CHOLESTA-5,24-DIEN-3BETA-OL
	547	58 999	0 398 24-METHYLCHOLESTA-5,22E-DIEN-3BETA-OL
23	999	111 999	0 384 CHOLESTA-5,24-DIEN-3BETA-OL
	631	72 999	0 384 (24S)-27-NOR-24-METHYLCHOLESTA-5,22E-DIE
	642	61 999	0 384 CHOLESTA-5,22E-DIEN-3BETA-OL
	581	68 999	0 426 GORGOST-5-EN-3BETA-OL
25	999	96 999	0 386 CHOLEST-5-EN-3BETA-OL
	647	55 999	0 386 5ALPHA-CHOLEST-7-EN-3BETA-OL
26	999	85 999	0 386 5ALPHA-CHOLEST-7-EN-3BETA-OL
	593	57 999	0 386 CHOLEST-5-FN-3BETA-OL
	581	50 999	0 400 5ALPHA-24-METHYLCHOLEST-7-FN-3BETA-OL
	500	37 999	0 426 (24Z)-24-PROPYLIDENECHOLESTA-5-FN-3BETA-
27	999	104 999	0 386 5ALPHA-CHOLESTAN-3BETA-OL
	629	68 999	0 388 5BETA-CHOLESTAN-3BETA-OL

STEROL		STEROLS MATCHED	
#	RANK	STEROL	
28	999	105 999	0 384 CHOLESTA-5,7-DIEN-3BETA-OL
30	999	82 999	0 372 19-NOR-CHOLEST-5-EN-3BETA-OL
32	999	108 999	0 388 5BETA-CHOLESTAN-3BETA-OL
	673	70 999	0 388 5ALPHA-CHOLESTAN-3BETA-OL
38	999	106 999	0 398 24-METHYLCHOLESTA-5,22E-DIEN-3BETA-OL
	669	79 999	0 398 5ALPHA-24-METHYLCHOLESTA-7,22E-3BETA-OL
	614	70 0	0 398 (24S)-24-METHYLCHOLESTA-5,25-DIEN-3BETA-OL
	581	68 999	0 426 GORGOST-5-EN-3BETA-OL
	556	59 999	0 398 24-METHYLCHOLESTA-5,24(28)-DIEN-3BETA-OL
41	999	104 999	0 396 24-METHYLCHOLESTA-5,7,22E-TRIEN-3BETA-OL
	558	48 0	0 410 24-ETHYLCHOLESTA-5,7,22E-TRIEN-3BETA-OL
43	999	114 0	0 398 (24S)-24-METHYLCHOLESTA-5,25-DIEN-3BETA-OL
	660	70 999	0 398 24-METHYLCHOLESTA-5,22E-DIEN-3BETA-OL
	635	75 999	0 398 5ALPHA-24-METHYLCHOLESTA-7,22E-3BETA-OL
	516	64 0	0 412 (24S)-24-ETHYLCHOLESTA-5,25-DIEN-3BETA-OL
44	999	106 999	0 398 24-METHYLCHOLESTA-5,24(28)-DIEN-3BETA-OL
	547	58 999	0 398 24-METHYLCHOLESTA-5,22E-DIEN-3BETA-OL
	521	61 999	0 426 GORGOST-5-EN-3BETA-OL
	527	39 999	0 426 (24Z)-24-PROPYLIDENECHOLESTA-5-EN-3BETA-OL
48	999	132 999	0 400 24-METHYLCHOLEST-5-EN-3BETA-OL
49	999	86 999	0 400 5ALPHA-24-METHYLCHOLEST-7-EN-3BETA-OL
	666	56 999	0 414 5ALPHA-24-ETHYLCHOLEST-7-EN-3BETA-OL
	529	45 999	0 386 5ALPHA-CHOLEST-7-EN-3BETA-OL
54	999	107 999	0 402 5ALPHA-24-METHYLCHOLESTAN-3BETA-OL
	584	59 999	0 416 5ALPHA-24-ETHYLCHOLESTAN-3BETA-OL
60	999	115 999	0 412 24-ETHYLCHOLESTA-5,22E-DIEN-3BETA-OL
	552	68 0	0 412 (25S)-24,27-DIMETHYLCHOLESTA-5,24(28)-DIEN-3BETA-OL
	547	46 999	0 412 23,24-DIMETHYLCHOLESTA-5,22-DIEN-3BETA-OL
	522	58 999	0 412 5ALPHA-24-ETHYLCHOLESTA-7,22E-DIEN-3BETA-OL
61	999	111 999	0 412 5ALPHA-24-ETHYLCHOLESTA-7,22E-DIEN-3BETA-OL
	529	62 0	0 412 23-NOR-GORGOST-5-EN-3BETA-OL
	513	59 999	0 412 24-ETHYLCHOLESTA-5,22E-DIEN-3BETA-OL
	523	44 999	0 412 23,24-DIMETHYLCHOLESTA-5,22-DIEN-3BETA-OL
63	999	86 0	0 410 24-ETHYLCHOLESTA-5,7,22E-TRIEN-3BETA-OL
	519	54 999	0 396 24-METHYLCHOLESTA-5,7,22E-TRIEN-3BETA-OL
68	999	123 999	0 412 24-ETHYLCHOLESTA-5,24(28)E-DIEN-3BETA-OL
	705	67 0	0 412 24-ETHYLCHOLESTA-5,24(28)Z-DIEN-3BETA-OL
	648	48 999	0 426 (24Z)-24-PROPYLIDENECHOLESTA-5-EN-3BETA-OL
	588	73 0	0 412 (24S)-24-ETHYLCHOLESTA-5,25-DIEN-3BETA-OL
	581	68 0	0 412 23-NOR-GORGOST-5-EN-3BETA-OL
71	999	95 0	0 412 24-ETHYLCHOLESTA-5,24(28)Z-DIEN-3BETA-OL
	626	77 999	0 412 24-ETHYLCHOLESTA-5,24(28)E-DIEN-3BETA-OL
	648	48 999	0 426 (24Z)-24-PROPYLIDENECHOLESTA-5-EN-3BETA-OL
	581	68 0	0 412 23-NOR-GORGOST-5-EN-3BETA-OL
	512	63 0	0 412 (25S)-24,27-DIMETHYLCHOLESTA-5,24(28)-DIEN-3BETA-OL

(cont.)

## STEROL

## STEROLS MATCHED

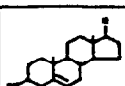
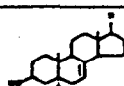
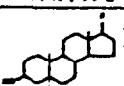
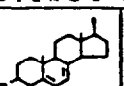
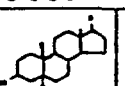
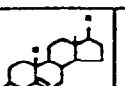
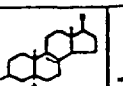
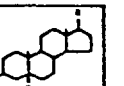






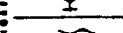


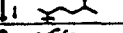



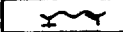






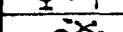
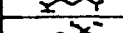

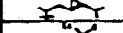
















#	RANK	STEROL
74	999 124 0	0 412 (24S)-24-ETHYLCHOLESTA-5,25-DIEN-3BETA-0
	585 72 999	0 412 24-ETHYLCHOLESTA-5,24(28)E-DIEN-3BETA-0L
	536 66 0	0 412 (25S)-24,27-DIMETHYLCHOLESTA-5,24(28)-DI
	504 59 0	0 412 23-NOR-GORGOST-5-EN-3BETA-0L
75	999 84 999	0 412 23,24-DIMETHYLCHOLESTA-5,22-DIEN-3BETA-0
77	999 123 0	0 412 (25S)-24,27-DIMETHYLCHOLESTA-5,24(28)-DI
	598 70 0	0 412 23-NOR-GORGOST-5-EN-3BETA-0L
	573 66 999	0 412 24-ETHYLCHOLESTA-5,22E-DIEN-3BETA-0L
	560 69 999	0 412 24-ETHYLCHOLESTA-5,24(28)E-DIEN-3BETA-0L
	548 68 999	0 412 (24S)-24-ETHYLCHOLESTA-5,25-DIEN-3BETA-0
78	999 117 0	0 412 23-NOR-GORGOST-5-EN-3BETA-0L
	621 59 999	0 412 24-ETHYLCHOLESTA-5,24(28)Z-DIEN-3BETA-0L
	569 70 999	0 412 (25S)-24,27-DIMETHYLCHOLESTA-5,24(28)-DI
	560 69 999	0 412 24-ETHYLCHOLESTA-5,24(28)E-DIEN-3BETA-0L
	516 64 999	0 412 (24S)-24-ETHYLCHOLESTA-5,25-DIEN-3BETA-0
79	999 103 999	0 414 5ALPHA-23-NOR-GORGOSTAN-3BETA-0L
80	999 132 999	0 414 24-ETHYLCHOLEST-5-EN-3BETA-0L
	716 91 0	0 0 (24R,27S)-24,27-DIMETHYLCHOLEST-5-EN-3BE
82	999 101 999	0 416 5ALPHA-24-ETHYLCHOLESTAN-3BETA-0L
	523 56 999	0 402 5ALPHA-24-METHYLCHOLESTAN-3BETA-0L
86	999 105 999	0 416 5BETA-24-ETHYLCHOLESTAN-3BETA-0L (S01533-
88	999 84 999	0 414 5ALPHA-24-ETHYLCHOLEST-7-EN-3BETA-0L
	569 49 999	0 406 5ALPHA-24-METHYLCHOLEST-7-EN-3BETA-0L
89	999 101 999	0 412 24-ETHYLCHOLESTA-5,7-DIEN-3-BETA-0L
90	999 127 0	0 0 (24R,27S)-24,27-DIMETHYLCHOLEST-5-EN-3BE
	734 97 999	0 414 24-ETHYLCHOLEST-5-EN-3BETA-0L
	500 42 999	0 414 5ALPHA-24-ETHYLCHOLEST-7-EN-3BETA-0L
91	999 117 999	0 426 GORGOST-5-EN-3BETA-0L
	743 55 999	0 426 (24Z)-24-PROPYLIDENECHOLESTA-5-EN-3BETA-
	560 51 999	0 426 5ALPHA-GORGOST-7-EN-3BETA-0L
	516 61 999	0 426 24-ISOPROPYLCHOLESTA-5,22E-3BETA-0L (S01
	513 57 999	0 384 CHOLESTA-5,24-DIEN-3BETA-0L
92	999 91 999	0 426 5ALPHA-GORGOST-7-EN-3BETA-0L
93	999 123 999	0 428 5ALPHA-GORGOSTAN-3BETA-0L
94	999 74 999	0 426 (24Z)-24-PROPYLIDENECHOLESTA-5-EN-3BETA-
	538 63 999	0 426 GORGOST-5-EN-3BETA-0L
	505 48 999	0 412 24-ETHYLCHOLESTA-5,24(28)Z-DIEN-3BETA-0L
96	999 118 999	0 426 24-ISOPROPYLCHOLESTA-5,22E-3BETA-0L (S01
	529 62 999	0 426 GORGOST-5-EN-3BETA-0L



An important limitation of the file search system is then its inability to distinguish between variations in side chain alkylation. These various side chain alkylation patterns are very important with respect to biosynthetic processes. Since these sterols have different retention indices, this limitation has been overcome by searching a file of retention indices as well as mass spectra. A computer program for accurately calculating retention indices has been developed by William Yeager, Department of Genetics, Stanford University, and is applicable to the rapid analysis sequence. Michael Kohraman has prepared a file of carefully measured retention indices from samples used to compile the mass spectral file; (Table VI) therefore, the limitations concerning identification of isomeric side chain alkylation patterns have been reduced.

See p. 89a for Table VI.

TABLE VI  
RETENTION INDICES OF STEROLS OF SP2250

MARINE STEROL NUCLEI									
MARINE STEROL SIDE CHAINS	NUCLEI								
									
C <sub>15</sub>		2650 1							
		2892 2							
									
		2824 4							
		2972 5							
		2971 6		2927 7					
		3047 8							
		3282 9	3268 10	3317 11		3091 12			
		3258 13							
		3411 14							
C <sub>16</sub>		3319 15	3385 16	3325 17					
		3328 18		19 20	21	3219 22			
		3442 23		24					
		3361 25	3334 26	3383 27	3402 28	3245 29	3290 30		31 3322 32
				33					
			34	35	3364 36	37			
		3420 38	3497 39	3480 40	3492 41	3292 42			
		3468 43							
		3473 44		45 46		47			
		3445 48	3538 49	3474 50	51				
C <sub>17</sub>			52	53	54	55	3348 56	57	58
			59						
		3492 60	3566 61		3557 63				
		3492 66	3624 67	3552 62		3376 64		65	
		3574 68		69 70					
		3601 71	3654 72	73					
		3555 74							
		3489 75					76		
		3568 77							
		3574 78		3560 79					
C <sub>18</sub>		3536 80		81 3560 82	83				
		3553 87		88	89	3427 84		85	86
		3578 90							
		3684 91	3755 92	3689 93					
			94						
			95						
		3487 96							
		3578 97							
									
									

Second, some sterols have very distinctive mass spectra with respect to the other spectra in the file, and no other spectrum ranks above 500 (for 17 spectra); however, the majority of spectra do show some similarities to other spectra in the file, i.e, have a cross rank > 500 with another sterol mass spectrum in the file. It is interesting that sterols which are saturated match only with other saturated sterols, sterols with one nuclear unsaturation match only with other sterols with one nuclear unsaturation, sterols with 2 nuclear unsaturations match only sterols with 2 nuclear unsaturations, and sterols with one nuclear and one side chain unsaturation (or ring junction) match only sterols possessing that property. The empirical ranking algorithm described previously has detected the number and general positions of unsaturation in the sterols. Therefore, if a new sterol is detected by the file search procedures then the general structural properties of the new sterol (number of nuclear and side chain double bonds) may be indicated by the structures of the sterols with which it is ranked even though the ranking values are very low.

The real utility of the search system will be in rapidly sorting a tremendous quantity of experimental data in an effort to reveal the sterols of novel structure. This is of tremendous utility because marine sterol mixtures are generally complex, containing over 40 sterols in some cases. However, once the sterol of novel structure is pointed out, then a careful analysis of the mass spectral fragmentation in terms of known processes must proceed. Rules generated via INTSUM, etc. analyses of the extensive marine sterol high resolution mass spectral files will help greatly by providing firm guidelines for the structural evaluations of the previously unencountered sterols.

#### 4.2.2 Researchers Receiving Marine Sterol Data

Dr. J. B. Heather  
The Upjohn Company  
Chemical Process,  
Rsch & Development  
Kalamazoo, Mich.

Dr. Steven C. Welch  
Dept of Chemistry  
University of Houston  
Houston, Texas 77004

Dr. Richard M. Wing  
Univ of California  
Riverside, Ca.

Prof. Paul J. Scheuer  
University of Hawaii  
2545 The Mall  
Dept of Chemistry  
Honolulu, Hawaii

Dr. Yuzura Shimizu  
Univ of Rhode Island  
College of Pharmacy  
53 Fogarty  
Kingston, R.I.

Dr. Maktoob Alam  
University of Houston  
College of Pharmacy  
Dept. of Med. Chem.  
and Pharmacognosy  
Houston, Texas 77004

Dr. Ron Quinn  
Roche Research Inst.  
P. O. Box 255  
Dee Why NSW 2099  
AUSTRALIA

Dr. K. Ivanetich  
Dept Physiol. & Med.  
Biochemistry  
Medical School  
Observatory, Cape  
SOUTH AFRICA

## 5 Carbon-13 Work

The work described in this section was accomplished in conjunction with work on structure elucidation and theory formation programs (sections 2 and 4). It is presented together here to make a more coherent presentation.

Carbon-13 nuclear magnetic resonance (CMR) has developed into an important tool for the structural chemist. A natural abundance CMR spectrum which is fully proton decoupled consists of a number of sharp peaks which correspond to the resonance frequencies in an applied magnetic field of the various types of carbon atoms present. A  $^{13}\text{C}$  shift is the amount an observed peak is shifted from that of a reference peak, usually tetramethylsilane (TMS).

In last year's annual report we discussed an extension of Meta-DENDRAL which allowed the program to form rules in the domain of CMR spectroscopy. During the past year we continued work on this program, and wrote a second program which applies CMR rules to structure elucidation problems. Rules generated from a combined set of paraffins and acyclic amines have been used to successfully identify the  $^{13}\text{C}$  NMR spectra of molecules not in the training set data. The introduction of a limited set of stereochemical terms to the rule generation procedure demonstrated the feasibility of extending the method to more complicated systems. A description of the rule formation and structure elucidation programs is given in [17]. Results are presented there for the combined set of paraffin and acyclic amines, as well as for a combined set of trans decalins and monohydroxylated androstanes.

### 5.1 Rule Formation Results

A set of rules was generated using a subset of the paraffin data from Lindeman and Adams<sup>12</sup> combined with a subset of the acyclic amine data from Eggert and Djerassi<sup>13</sup>. Molecules with the empirical formula  $\text{C}_9\text{H}_{20}$  and  $\text{C}_6\text{H}_{15}\text{N}$  were excluded from the training set for later use in testing the generality of the rules. The rule set was tested by generating all structural isomers with the empirical formulas  $\text{C}_9\text{H}_{20}$  (35 isomers) and  $\text{C}_6\text{H}_{15}\text{N}$  (39 isomers), predicting the spectrum of each isomer, then ranking the predicted spectra by similarity to a known spectrum. The rank of the predicted spectra associated with the correct candidate structure provides an indication of the utility and

---

<sup>12</sup> Lindeman, L.P. and J.Q. Adams, *Anal. Chem.*, (1971), 43,p. 1245.

<sup>13</sup> Eggert, H. and C. Djerassi, *J. Amer. Chem. Soc.* (1973), 95,p. 3710.

validity of the generated rules. For the above test we used the 24  $C_9H_{20}$  spectra available from the work of Lindeman and Adams. The predicted spectra of the 35 structural isomers were compared and ranked against each of these available spectra. The results of this ranking for  $C_9H_{20}$  as well as a similar test on  $C_6H_{15}N$  are shown in Table VII.

Empirical Formula	Number of Candidates	Number of Spectra	Rank of Correct Structure (Freq of Correct Ranking)				
			1 <sup>st</sup>	2 <sup>nd</sup>	...6 <sup>th</sup>	..9 <sup>th</sup>	
$C_9 H_{29}$	35	24	20/24	3/24			1/24
$C_6 H_{15} N$	39	11	8/11	2/11	1/11		

Table VII. Results of Structure Ranking

## 5.2 Adding Stereochemistry to the Rule Language

The work on the paraffins and acyclic amines requires only topological descriptors in the language of atom features. Because of the dependence of  $^{13}C$  shifts on stereochemical features<sup>14 15</sup> it is necessary to have the facility to include stereochemical terms when they are required. Substituents placed on systems which have static conformations such as trans decalin and androstane with trans ring fusions can be described in discrete terms. The terms we selected describe the orientation on the ring of the substituent as either axial or equatorial, and either alpha or beta. For instance, a substituent is beta in 10-methyl-trans-decalin if it is on the same side of the ring as the methyl group and alpha if on the opposite side of the ring from the methyl group. The rule generation program with the extension of the language to include these atom features was run on a combined set of trans decalins, 10-methyl-trans-decalols and monohydroxylated androstanes with trans ring fusions selected from the works of Grover and Stothers<sup>16</sup> and Eggert et. al.<sup>17</sup>

<sup>14</sup> Grover, S.H. and J.B. Stothers, Can. J. Chem. (1974), 52, p. 870.

<sup>15</sup> Eggert, H., C. VanAntwerp, N. Bhacca, and C. Djerassi, J. Org. Chem., (1976), 41, p. 71.

<sup>16</sup> Grover, Op. cit.

<sup>17</sup> Eggert, Op. cit.

Sixty rules were generated to cover the 249 data peaks of 17 compounds. Samples of the rules generated are shown in Figure 11. The examination of these rules will show that they are useful for the chemist who wants to study contributions to the total shift as well as for structure elucidation.

See p. 94a for rules.

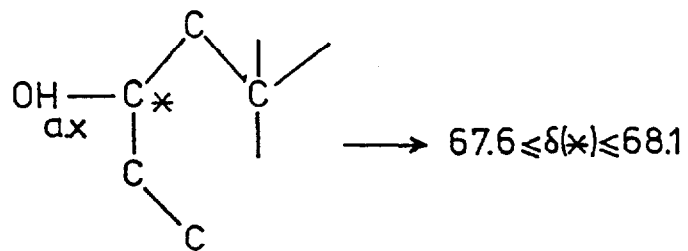
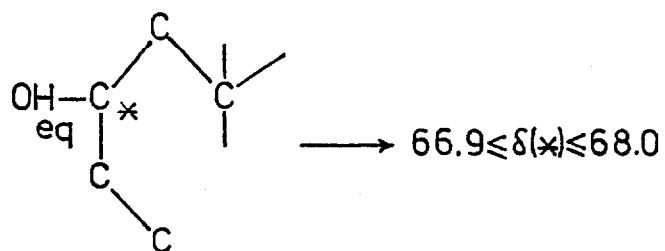
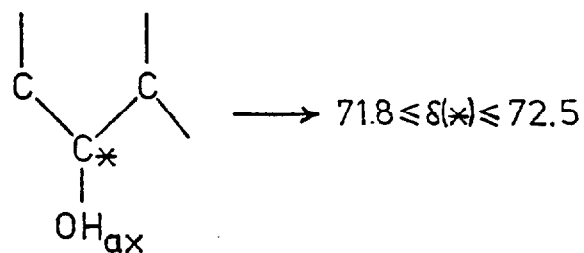
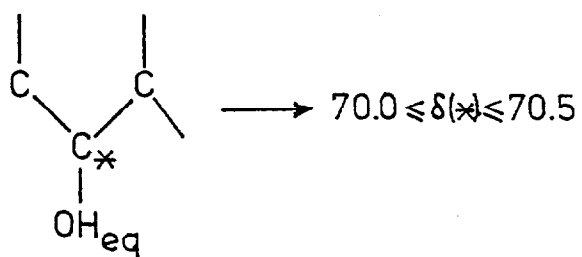
Figure 11. Sample rules constructed from decalins and hydroxy steroids with trans ring fusions. The '\*' identifies the carbon atom to which the shift is assigned. is in ppm downfield from TMS.

### 5.3 Structure Elucidation

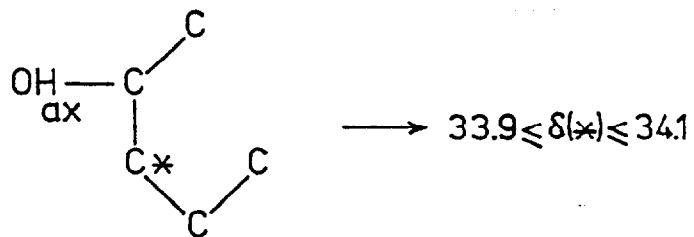
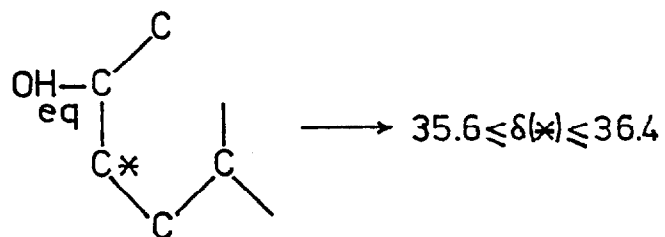
Molecular structure elucidation using CMR consists of using a set of rules which summarize the CMR behavior of a set of compounds to identify other unknown compounds within that or similar classes. The information which the chemist must supply to the structure elucidation program includes the empirical formula of the unknown as well as its observed spectrum. Two parameters may be set by the chemist to select the number of plausible structures to be determined, and to specify the error range in ppm which should be assigned to the rules to account for deficiencies in the training data, experimental error, solvent effects, etc. From this information and its store of CMR rules, the program assembles a set of structures which are plausible sources of the unknown spectrum.

Molecular structure elucidation is accomplished by our program by selecting a shift (peak) in the observed spectrum, then finding the rules which are possible explanations for this shift. The rules selected postulate partial substructures which

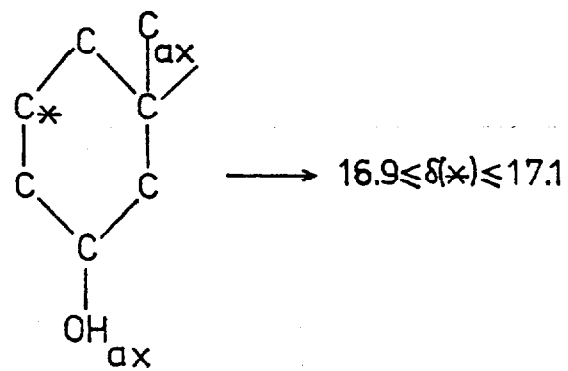
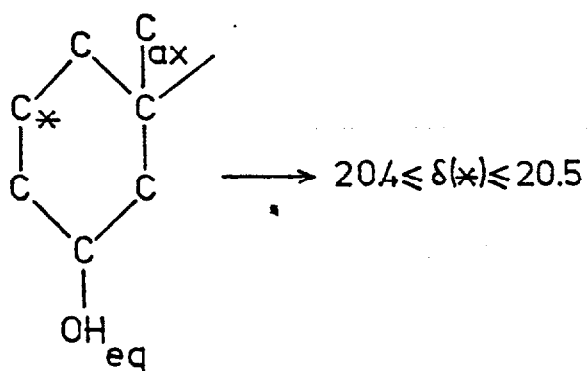
## Alpha Carbon Rules



## Beta Carbon Rules



## Gamma Carbon Rules





might be in the molecule. These substructures are then assembled jigsaw puzzle fashion to construct the final molecule. Constraints stemming from both the observed spectrum and information associated with each rule are used to constrain the process so that only "reasonable" structures will be considered.

The structure elucidation program has been run on several test cases using unknown paraffin and acyclic amine spectra with reasonable success. This program is described in detail in [17].

#### 5.4 Geometric Distortions in Steroids

For a given molecule, deviations between its observed  $^{13}\text{C}$  NMR spectrum and its spectrum predicted from a set of empirical  $^{13}\text{C}$  NMR rules is often explained in terms of geometric distortions. In order to examine the effect of geometric distortions on  $^{13}\text{C}$  NMR shifts, Allinger's<sup>18, 19</sup> molecular force field program has been used to model geometric distortions in mono-hydroxy- 5 alpha, 14-alpha androstanes. The net effect of many types of slight geometric distortions on the  $^{13}\text{C}$  shift were examined in terms of the non-bonded interactions. The delta(alpha) and delta(beta) effects could be characterized in a few terms suggested by the non-bonded interactions. The results of this study were published in [16].

### 6 DATA COLLECTION AND DATA REDUCTION

#### 6.1 DENDRAL GC/MS and MS Work

The following is a summary of the activities in the GC/MS lab for the past year. This work involves both development of the GC/MS Computer systems for both high and low resolution GC/MS applications and application of the existing system to mass spectral analyses of compounds of biomedical importance.

##### A) Low resolution GC/MS: (manual mode)

93 sterol mixtures (marine sterol extractions) for Dr. Djerassi's group. Identification of free sterols.

##### B) High Resolution GC/MS

---

<sup>18</sup> N.L. Allinger, M.T. Tribble, M.A. Miller and D.H. Wertz, J. Amer. Chem. Soc., 93, 1637 (1971).

<sup>19</sup> D.H. Wertz and N.L. Allinger, Tetrahedron, 30, 1579 (1974).

## Total sample mixtures: 86

for:	1) Dr. Djerassi	52
	2) Genetics	25
	3) Prof. Adlercreutz, Finland	9

1) Dr. Djerassi: all marine sterols, especially for library purposes and thesis of Bob Carlson.

2) Genetics: Urine extractions, channel-black and carbon-black, all for assistance in identification of unknown compounds whose structures could not be elucidated by low resolution mass spectral data coupled with library search.

3) Prof. Adlercreutz, Clinical Chemistry, University of Helsinki, Finland needed quantitation of a corticosteroid. Tests were made to find sensitivity limit with Aldosterone-TMS. 5 ug Aldosterone-TMS was limitation. An unknown corticosteroid with a M+504 (low resolution spectrum) could not be identified by H. R. GC/MS due to amount of sample availability plus lack of sensitivity on our instrument. The sample was a substance occurring in patients who have no aldosterone, but still may have hypertension or hypokalemia.

## High resolution MS

## 43 samples total:

for:	1) Dr. Djerassi	29
	2) Prof. Fringuelli, Italia	8
	3) Prof. Nakano, Venezuela	6

1) Dr. Djerassi: Structure identification of new sterols plus terpenes.

2) Prof. Fringuelli, Perugia University, Perugia, Italia. Had 8 samples of furan, thiophen, selenophen and tellurophen derivatives for mass fragmentation studies. H. R. resolved all isotopes of each substance (up to 8 isotopes) and gave clear identification pattern. He is preparing and sending us more sets of compounds.

3) Prof. Nakano, Instituto Venezolano, Caracas, Venezuela, needed high resolution spectra of Oxadiazole derivatives for fragmentation studies, and successful identification of all six samples were possible.

## Computerized MS (Incl. trials)

H. R. (R-10000) + GC/MS H.R. R-5000

Start Jan. 77.		Total
S0 1696 to 1921	DOS (*)	225
S0 1839 to 1859	DOS duplication nos.	20

SO 1960 to 2379	RT-11	419
SO 1883 to 1955	DOS	72
SO 2437 to 2477	RT-11	40
SO 1956 to 2031	DOS	75
SO 2479 to 2481	RT-11	2
SO 2032 to 2037	DOS	5
SO 2516 to 2907	RT-11	391

Total samples tested	1249
----------------------	------

\* DOS and RT-11 refer to the two different operating systems for the PDP-11 computer system. During the past year we have had to convert operating systems.

## 6.2 Collaborators Receiving the CLEANUP and HISLIB Programs

Following is an alphabetical list of people who have requested copies of the program for extracting better resolved mass spectra from GC/MS data, described in [10].

Dr. Craig Anderson  
Gulf South Research Institute  
P.O. Box 26500  
New Orleans, Louisiana 70186

Dr. John B. Bagger  
Department of Chemistry  
Colorado State University  
Fort Collins, Colorado 80521

Dr. Rod Britten  
Jet Propulsion Laboratories  
4800 Oak Grove Drive, 168-227  
Pasadena, California 91103

Dr. Robert D. Brown  
Bristol Laboratories  
P. O. Box 657  
Syracuse, New York 13201

Dr. Peter Bruck  
Magyar Tudományos Akademia  
Kozponti Kemiai Kutato Intezete  
1088 Budapest Puskin u. 11-13.  
Hungary

Dr. Lawrence Burkhard  
Water Chemistry Laboratory  
University of Wisconsin  
660 North Park Street  
Madison, Wisconsin 53706

Dr. Richard M. Caprioli  
Dr. William E. Seifert, Jr.  
Program in Biomolecular Analysis  
Univ of Texas Medical School  
P. O. Box 20708  
Houston, Texas 77025

Dr. Henry E. Dayringer  
Mail Zone V1A  
Monsanto Agricultural Research  
800 North Lindbergh Boulevard  
St. Louis, Missouri 63166

Dr. James F. Elder  
574 Building  
Analytical Laboratories  
Dow Chemical U.S.A.  
Midland, Michigan 48640

Dr. W.K. Elkin  
Department of Toxicology  
Swedish Medical Research Council  
Karolinska Institutet  
S-104 01 Stockholm, Sweden

Dr. Paul V. Fennessey  
B.F. Stolinsky Rsch Laboratories  
Department of Pediatrics  
Univ of Colorado Medical Ctr  
4200 East Ninth Avenue  
Denver, Colorado 80220

Dr. Claude Finn  
School of Pharmacy  
U. C. San Francisco Medical Ctr  
San Francisco, California 94143

Dr. R. Fluckiger  
Balzers Aktiengesellschaft fur  
Hochvakuumtechnik und Dunne  
Schichten  
FL-9496 Balzers  
Furstentum Liechtenstein

Dr. A.N. Freedman  
Central Electricity  
Research Laboratories  
Kelvin Avenue, Leatherhead  
Surrey, England

Dr. Nelson M. Frew  
Chemistry Department  
Woods Hole Oceanographic  
Institution  
Woods Hole, Massachusetts 02543

Dr. Richard Gans  
Chemical Research Division  
Bound Brook Laboratories  
American Cyanamid Company  
Bound Brook, New Jersey 08805

Mrs. E.M. Gomm  
Department of Chemistry  
University of Natal  
P.O. Box 375, Pietermaritzburg  
Natal, South Africa

Dr. Sydney M. Gordon  
Chemistry Division  
Atomic Energy Board  
Private Bag 256, Pretoria  
Republic of South Africa

Dr. Richard A. Graham  
FSL  
U. S. Army Natick Laboratories  
Natick, Massachusetts 01760

Dr. H. Knoepfel  
EURATOM - CCR  
Casella Postale No. 1  
Ispra, Italy

Dr. Donald A. Griffin  
Dept of Agricultural Chemistry  
Oregon State University  
Corvallis, Oregon 97331

Dr. G. Knowles  
Water Research Centre  
Stevenage Laboratory  
Elder Way, Stevenage  
Hertfordshire SG1 1TH, England

Dr. William Haddon  
Western Regional Research Center  
U.S. Department of Agriculture  
800 Buchanan Street  
Albany, California 94710

Dr. Thomas Knudsen  
Northrop Services  
Box 12313  
Research Triangle Park, N.C.

Dr. P.T. Holland  
Ministry of Agriculture  
and Fisheries  
Private Bag, Hamilton  
New Zealand

Dr. Douglas W. Kuehl  
Mass Spectrometry Laboratory  
Environmental Rsch Laboratory  
6201 Congdon Boulevard  
Duluth, Minnesota 55804

Dr. I. Howe  
Shell Biosciences Laboratory  
Sittingbourne Research Centre  
Sittingbourne, Kent ME9 8AG,  
England

Dr. Ake Lundin  
LKB-PRODUKTER AB  
Molecular Analysis Division  
S-161 25 Bromma 1  
Sweden

Akio Ide, Ph.D.  
Ehime University  
Agricultural Chemistry Dept  
Matsuyama 790, Japan

Dr. John L. MacDonald  
Central Research  
Ralston Purina Company  
Checkerboard Square  
St. Louis, Missouri 63188

Dr. J. B. Justice  
Emory University  
Atlanta, Georgia 30322

Dr. R.G.A.R. MacLagan  
Department of Chemistry  
University of Canterbury  
Christchurch 1, New Zealand

Dr. Graham S. King  
Department of Chemical Pathology  
Queen Charlotte's Hospital  
Goldhawk Road  
London, ENGLAND W6 0XG

Dr. John C. Marshall  
Department of Chemistry  
The University of North Carolina  
Chapel Hill, N.C.

Dr. Daniel R. Knapp  
Department of Pharmacology  
Medical Univ of South Carolina  
80 Barre Street  
Charleston, South Carolina 29401

Dr. R. A. F. Matheson  
Chemistry Section  
Environmental Protection Service  
5151 George Street  
Halifax, Nova Scotia CANADA

Dr. James A. McCloskey, Jr. Professor, Biomedical Chemistry Dept. Biopharmaceutical Sciences University of Utah Salt Lake City, Utah 84112	Dr. Philip W. Ryan Battelle Pacific Northwest Laboratories, 329 Bldg. Battelle Boulevard Richland, Washington 99352
Dr. Ingolf Meineke Fachbereich Chemie Philipps Universitaet 3550 Marburg/Lahn, Lahnberge WEST GERMANY	Dr. Robert S. Schroeder Gulf Oil Chemicals Company P. O. Box 2900 Shawnee Mission, Kansas 66201
Dr. Roy O. Morris Dept. Agricultural Chemistry Oregon State University Corvallis, Oregon 97331	Dr. J. Scrivens Imperial Chemical Industries PO Box 90 Wilton Middlesbrough Cleveland TS6 8JE England
Dr. James E. Oberholtzer Arthur D. Little, Inc. Acorn Park Cambridge, Massachusetts 02140	Dr. Walter M. Shackelford Analytical Chemistry Branch Environmental Rsch Laboratory Athens, Georgia 30601
Mr. Andrew Pallos Aerospace Corporation P.O. Box 92957 Los Angeles, California 90009	Dr. M.A. Shaw Unilever Research Port Sunlight Laboratory Port Sunlight Wirral, Merseyside L62 4XN, England
Mr. Dan Pearce Orange Co Sheriff-Coroner Dept 550 N. Flower Street Santa Ana, California 92702	Dr. Jacob Shen The Standard Oil Company 4440 Warrensville Center Road Cleveland, Ohio 44128
Dr. William R. Penrose Newfoundland Biological Station 3 Water St. East St. John's, Newfoundland A1C 1A1	Dr. M. M. Siegel FMC Corporation Chemical Group Box 8 Princeton, New Jersey 08540
Dr. Ronald D. Plattner Northern Regional Research Lab. U.S. Department of Agriculture Peoria, Illinois 61604	Dr. G. P. Slater National Rsch Council of Canada Prairie Regional Laboratory 110 Gymnasium Road, University Campus Saskatoon, Saskatchewan CANADA
Ken Pocek Scientific Instruments Division Hewlett-Packard Company 1601 California Avenue Palo Alto, California 94304	Dr. Carroll A. Smith Div of Chemical Oceanography University of Miami 4600 Rickenbacker Causeway Miami, Florida 33149